Bioestadística

Sesión 3-4: Probabilidad

José Aurelio Pina Romero
Bioestadística – Grado Enfermería
UA- Departamento de Enfermería

Universitat d'Alacant
Universidad de Alicante

Sesión 1-2

- Propiedades básicas
  - $0 \leq P(A) \leq 1$
  - $P(\overline{A}) = 1 - P(A) \rightarrow P(A) + P(\overline{A}) = 1$
  - $P(A \cup B) = P(A) + P(B)$ si $A \cap B = \emptyset$
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ Cuando AB no son incompatibles

- Probabilidad condicionada

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B/A) = \frac{P(B \cap A)}{P(A)}$$

$$P(A \cap B) = P(B) \cdot P(A/B)$$

$$P(A \cap B) = P(A) \cdot P(B/A)$$

(IV) Los datos sobre el estudio sobre el hábito de lectura en estudiantes universitarios según género que ha puesto en marcha la Rectora de la UA son los siguientes

Género	<1 hora a la semana	Entre 1 y 2 h a la semana	> 2 h a la semana
Masculino	40	10	8
Femenino	20	20	22

Si seleccionamos uno de los sujetos al azar, conteste a las siguientes preguntas:

1. La probabilidad de ser género masculino y leer menos de 1 hora la semana es de 40/62 = 0,64

2. La probabilidad de leer menos de una hora a la semana es de 60/120=0,5

3. La probabilidad de que lea más de 2 horas a la semana si es género masculino es 8/120 = 0,067

4. La probabilidad de leer entre 1 y 2 h a la semana o de género femenino es 72/120=0,6

Leyes de Morgan

$$P\left(\overline{A}\cup\overline{B}\right)=P\left(\overline{A\cap B}\right)=1-P\left(A\cap B\right)$$

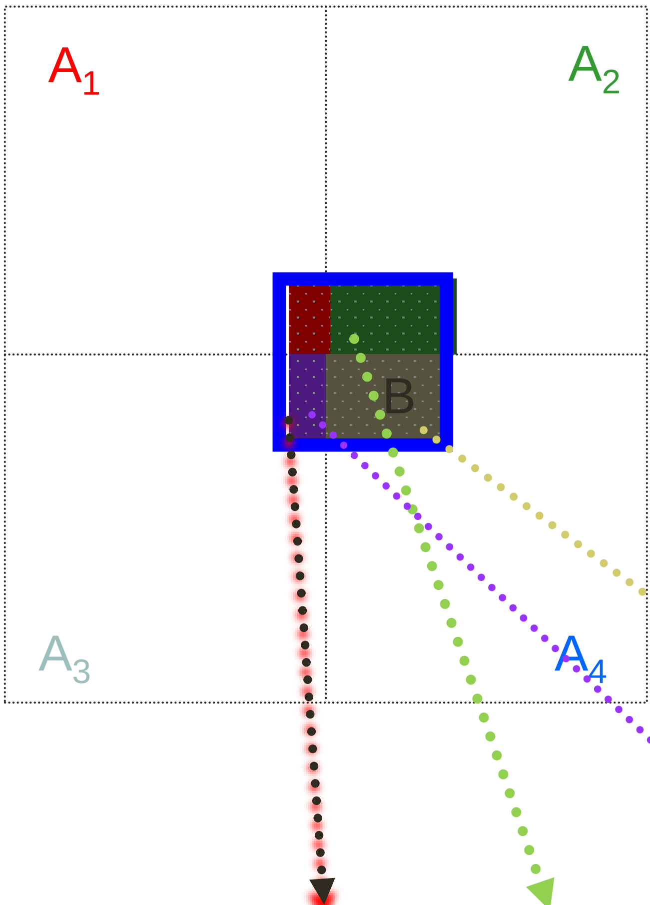
$$P\left(\overline{A}\cap\overline{B}\right)=P\left(\overline{A\cup B}\right)=1-P\left(A\cup B\right)$$

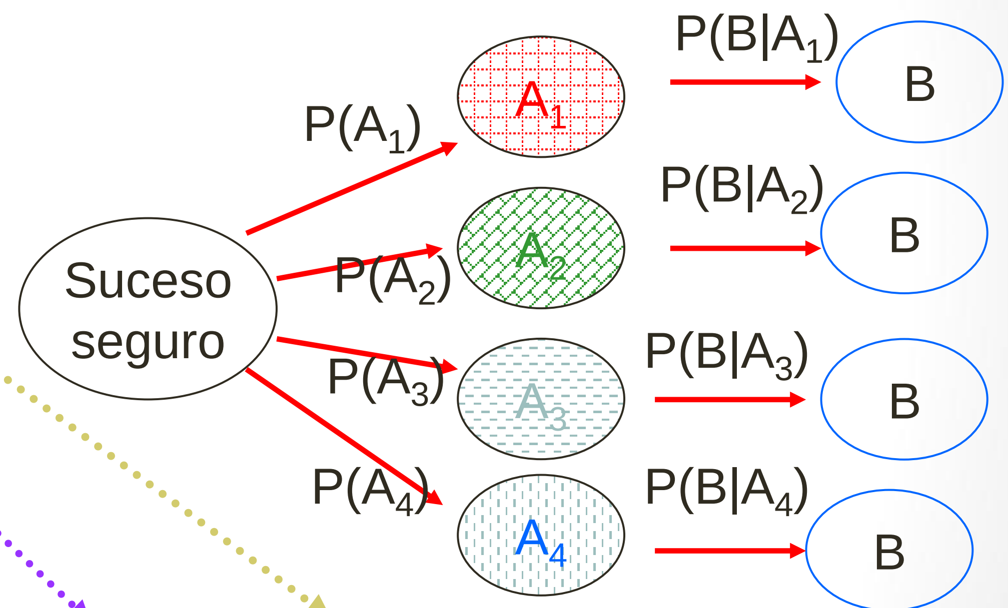
https://www.geogebra.org/m/t7s4dhhj

Teorema de la probabilidad total

Si conocemos la probabilidad de B en cada uno de los componentes de un sistema exhaustivo y excluyente de sucesos, entonces…

… podemos calcular la probabilidad de B.

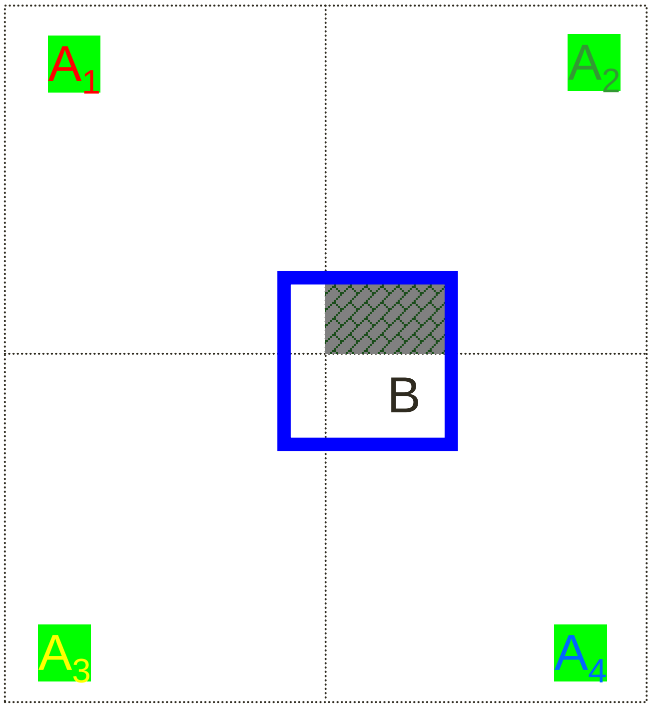




$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) + P(A_4 \cap B)$

$= P(A_1)\,P(B|A_1) + P(A_2)\,P(B|A_2) + \ldots = P(B) = \sum_{i=1}^{n} P(A_i \cap B) = \sum_{i=1}^{n} P(A_i) \cdot P(B/A_i)$

Teorema de Bayes



…si ocurre B, podemos calcular la probabilidad (*a posteriori*) de ocurrencia de cada A_i.

$$P(A_i|B)=\frac{P(A_i \cap B)}{P(B)}$$

donde P(B) se puede calcular usando el teorema de la probabilidad total:

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) + P(A_4 \cap B)$$

$$=P(A_1)\,P(B|A_1) + P(A_2)\,P(B|A_2) + \ldots =$$

MÁS EJEMPLOS

Ejemplo (I): En esta aula el 70% de los alumnos son mujeres. De ellas el 10% son fumadoras. De los hombres, son fumadores el 20%.

- Probabilidad de ser fumador
- Probabilidad de no ser fumador.
- Probabilidad de ser fumador o mujer.
- Probabilidad de ser fumador u hombre.
- Probabilidad de no ser fumador ni mujer.
- Probabilidad de ser fumador si es hombre.

Ejemplo (II): Suponga que un cierto tipo de alergia respiratoria generalmente afecta a 1 de cada 20 personas, mientras que la intolerancia alimentaria afecta al 3.5% de las personas. Suponiendo que los dos eventos son independientes,

¿Cuál es la probabilidad de no tener alergia respiratoria?

¿Cuál es la probabilidad de tener ambos problemas? y de tener solo uno?

Ejemplo (III):En un estudio sobre el índice de masa corporal (IMC) en una determinada población, se estimó que el 33% de las personas tenía un peso normal, el 50% tenía sobrepeso y el 17% era obeso. En estos 3 grupos, la probabilidad de aparición de cierto tipo de enfermedad cardiovascular es respectivamente del 1%, 3% y 6%. Suponiendo que el tamaño de la población es de 10 000 personas, calcule:

- ¿Cuántas personas obesas había en la población?
- ¿Cuál es la probabilidad de que un individuo elegido al azar de la población tenga un peso normal y se vea afectado por ese tipo de enfermedad cardiovascular?
- ¿cuántas personas enfermas (con enfermedad cardiovascular) esperamos tener en la población?

Pruebas diagnósticas

- Se entiende como prueba diagnóstica para una determinada enfermedad, al conjunto de intervenciones sobre un individuo encaminadas a “determinar” la presencia o no de enfermedad o su grado.

- Objetivo: debe de ser la reducción de incertidumbre sobre el estado de la enfermedad de la persona.

- Ideal: una prueba diagnóstica sería exacta si de la información que proporciona (resultados de la prueba) se desprende con exactitud el estado de la enfermedad, desaparición de la incertidumbre inicial.

Pruebas diagnósticas

Ejemplos

Prueba para detectar la presencia de un agente contaminante de un alimento.

Presencia de cierto grado de azúcar de un melón.

Cuestionario para diagnostica la presencia de un desorden alimenticio.

Prueba de Orina.

Test de embrazo.

Test de antígenos

Pruebas diagnósticas

Ejemplos

Ecografía diagnóstica de cáncer gástrico.

Test detectar VIH

Pruebas diagnósticas

	Resultados de la enfermedad	
Resultados de la prueba	E	$\overline{E}$ = S
+	a	b
-	c	d

a = número pacientes con la enfermedad diagnosticados como positivos por la prueba.

b = número de pacientes sin la enfermedad diagnosticados como positivos por la prueba.

c = número de pacientes sin la enfermedad diagnosticados como negativos por la prueba.

d = número de pacientes sin la enfermedad diagnosticados como “negativos” por la prueba.

Pruebas diagnósticas

	Resultados de la enfermedad	
Resultados de la prueba	E	$\overline{E}$ = S
+	a	b
-	c	d

Sensibilidad = P(+ | Enfermo)= a/(a+c)
Verdaderos positivos

Especificidad = P(- | Sano) = d / (b+d)
Verdaderos negativos

Pruebas diagnósticas

	Resultados de la enfermedad	
Resultados de la prueba	E	$\overline{E}$ =S
+	a	b
-	c	d

Falsos negativos = P(- | E) = c / (a+c)

Falsos positivos = P(+ I S) = b / (b+d)

Pruebas diagnósticas

	Resultados de la enfermedad	
Resultados de la prueba	E	$\overline{E}$ =S
+	a	b
-	c	d

S = Sensibilidad = P(+ | Enfermo)= a/(a+c)
E = Especificidad = P(- | Sano) = d / (b+d)

FN = Falsos negativos = P(- | E) = c / (a+c)
FP = Falsos positivos = P(+ I S) = b / (b+d)

S+ FN = ¿?
E+FP = ¿?

Pruebas diagnósticas

	Resultados de la enfermedad	
Resultados de la prueba	E	$\overline{E}$ = S
+	a	b
-	c	d

A partir de lo anterior y usando el teorema de Bayes, podemos calcular las probabilidades *a posteriori* (en función de los resultados del test): Valores predictivos

Valor predictivo positivo(VPP)=P(Enfermo|+)
Valor predictivo negativo(VPN) = P(Sano | -)

Pruebas diagnósticas

$$VPP = P\left(E / +\right) = \frac{P\left(E \cap +\right)}{P(+)} = \frac{P\left(+\middle|E\right)P\left(E\right)}{P\left(+\middle|E\right)P\left(E\right) + P\left(+\middle|\overline{E}\right)P\left(\overline{E}\right)} =$$

$$VPP = \frac{\text{Sensibilidad} \times \text{Prevalencia}}{\text{Sensibilidad} \times \text{Prevalencia} + (1-\text{Especificidad}) \times (1-\text{Prevalencia})} =$$

$$VPN = P\left(\overline{E} / -\right) = \frac{P\left(\overline{E} \cap -\right)}{P(-)} = \frac{P\left(-\middle|\overline{E}\right)P\left(\overline{E}\right)}{P\left(-\middle|E\right)P\left(E\right) + P\left(-\middle|\overline{E}\right)P\left(\overline{E}\right)} =$$

$$VPN = \frac{\text{Especificidad} \times (1\text{-Prevalencia})}{\text{Especificidad} \times (1\text{-Prevalencia}) + (1-\text{Sensibilidad}) \times \text{Prevalencia}} =$$

Ha sido ensayada una nueva prueba para la detección de diabetes. La prueba ha producido 138 resultados positivos sobre 150 personas de las que se sabía que eran diabéticos, mientras que sobre otras 150 personas no diabéticas ha producido también 24 resultados positivos.

i. Calcule la sensibilidad, especificidad y falsos positivos y negativos de la prueba.

ii. Si aplicamos esta prueba en un programa de detección precoz sobre una población con prevalencia de diabetes de 0.02 (2%), ¿cómo se llama y cuánto vale la probabilidad de que una persona sobre la que la prueba resulta positiva esté realmente enferma?

	E	$\overline{\mathbf{E}}$
+		
-		

Sensibilidad = P(+|E) =

Especificad = P(-| $\bar{E}$) =

Falsos POSITIVOS = P(+| $\bar{E}$) =

Falsos NEGATIVOS = P(-|E) =

ii.Nos dicen que la prevalencia es de 0,02 (P(Enf)=0,02) , lo que nos pide la pregunta es que calculemos el Valor Predictivo Positivo:

VP+= P(Enf|+)= $$\frac{P(Enf \cap +)}{P(+)} = \frac{P(+|Enf)P(Enf)}{P(+|Enf)P(Enf) + P(+|\overline{Enf})P(\overline{Enf})} =$$

$$= \frac{\text{Sensibilidad´ Prevalencia}}{\text{Sensibilidad´ Prevalencia} + (1 - \text{Especificidad})´\ (1 - \text{Prevalencia})} =$$

https://www.geogebra.org/m/gxwazpct

(II) En un laboratorio, han desarrollado una nueva prueba de diagnóstico para la enfermedad celíaca basada en una muestra de sangre. Necesitan precisión de la prueba mediante un experimento controlado. Toman 20 pacientes con enfermedad celíaca comprobada por biopsia ("enfermos") y 60 pacientes que fueron revisados debido a síntomas pero se descubrió que no tenían la enfermedad celíaca ("no enfermos"). Los 80 sujetos reciben la nueva prueba, los resultados son los siguientes:

	E	$\bar{\mathbf{E}}$
+	17	5
-	3	55

¿Cuál es la probabilidad de que un individuo con la enfermedad tenga una prueba positiva? ¿Y prueba negativa?

¿Cuál es la probabilidad de que un individuo sin la enfermedad tenga una prueba positiva? ¿Y prueba negativa?

(II) En un laboratorio, han desarrollado una nueva prueba de diagnóstico para la enfermedad celíaca basada en una muestra de sangre. Necesitan precisión de la prueba mediante un experimento controlado. Toman 20 pacientes con enfermedad celíaca comprobada por biopsia ("enfermos") y 60 pacientes que fueron revisados debido a síntomas pero se descubrió que no tenían la enfermedad celíaca ("no enfermos"). Los 80 sujetos reciben la nueva prueba, los resultados son los siguientes:

	E	**$\overline{E}$**
+	17	5
-	3	55

Ahora, van a aplicar esta nueva prueba en una población donde el 8% tiene esta enfermedad celíaca. En una persona seleccionada al azar computa.

La probabilidad de un resultado positivo y negativo de la prueba.

Si el resultado fue positivo, ¿cuál es la probabilidad de estar enfermo?

Si el resultado fue negativo, ¿cuál es la probabilidad de estar sano?

(III) El cáncer de colon es una de las principales causas de muerte por cáncer. Su diagnóstico exige la exploración de la totalidad del marco colónico mediante técnicas radiológicas o endoscópicas. En un estudio sobre el uso de la ecografía abdominal como prueba diagnóstica para la detección de este tipo de cáncer, se pudo determinar que la sensibilidad fue de 0,88 y la especificidad de 0,92.
A partir de estos datos:

i. Calcular la probabilidad de falso positivo y falso negativo.

ii. Si la prueba es aplicada sobre una población con prevalencia de cáncer de colon del 4% ¿cuáles son los valores predictivos (positivo y negativo) de la prueba?

iii. ¿Cuál será el valor predictivo positivo si la prueba se aplica sobre una población con prevalencia de cáncer de colon del 1%?

(III) En la siguiente tabla se muestran los resultados de un estudio para evaluar la utilidad de una tira reactiva para el diagnóstico de infección urinaria. Si se utiliza esta prueba diagnóstica sobre una población con una prevalencia de infección urinaria del 4%, calcule:

Tina reactiva	**I**	**$\bar{I}$**
Positiva	57	211
Negativa	7	262

a) Calcula la sensibilidad y especificidad
b) Calcula VPP y VPN

CUESTIONES - V/F

1) El falso positivo de una prueba diagnóstica es la probabilidad de que la prueba proporciones un resultado positivo sobre un individuo sano.
2) Falso positivo = 1- Especificidad
3) Falso negativo = 1- Sensibilidad
4) Falso positivo = 1 – Sensibilidad
5) Falso negativo = 1 – Especificidad
6) A mayor prevalencia de una patología en una determinada población, mayor valor predictivo positivo.
7) A menor prevalencia de una patología en una determinada población, mayor valor predictivo negativo.
8) Si $A \cap B \neq \emptyset \rightarrow P(A \cap B) = 0$